Differential item functioning for boys and girls in a screening instrument for attention deficit hyperactivity disorder

Appelbaum, S ; Lefering, R ; Wolff, C ; Tomasik, Martin J ; Ostermann, T

Abstract: Differential item functioning (DIF) indicates differential response probabilities of items for different subgroups. While there is a vast amount of research and literature on DIF in the field of educational screening and career assessment, DIF analysis has hardly been applied in the field of clinical assessment. This paper aims at analyzing the presence of gender related DIF in a crosssectional survey of children assessed by a structured questionnaire containing items on attention deficit and hyperactivity. A total of 1449 children (mean age: 1.94 +- 0.14 years; 51.2% male) were included. Almost no significant variations in parameters were found between boys and girls. Results based on a Partial Credit Model indicate an absence of DIF in eight out of nine items. Consistent with other studies in attention deficit hyperactivity disorder (ADHD) our results imply that the same level of rating for a symptom has the same meaning for boys and girls.

DOI: https://doi.org/10.3233/SHTI190797

German Medical Data Sciences: Shaping Change – Creative Solutions for Innovative Medicine
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI190797

Differential Item Functioning for Boys and Girls in a Screening Instrument for Attention Deficit Hyperactivity Disorder

Sebastian APPELBAUM[a], Rolf LEFERING[b], Christian WOLFF[c], Martin J. TOMASIK[a], Thomas OSTERMANN[a,1]
[a]Department for Psychology and Psychotherapy, Witten/Herdecke University, Germany
[b]Institute for Reasearch in Operative Medicine (IFOM), Witten/Herdecke University, Cologne, Germany
[c] Pediatric Practice, Hagen

Abstract. Differential item functioning (DIF) indicates differential response probabilities of items for different subgroups. While there is a vast amount of research and literature on DIF in the field of educational screening and career assessment, DIF analysis has hardly been applied in the field of clinical assessment. This paper aims at analyzing the presence of gender related DIF in a cross-sectional survey of children assessed by a structured questionnaire containing items on attention deficit and hyperactivity. A total of 1449 children (mean age: 1.94 ± 0.14 years; 51.2% male) were included. Almost no significant variations in parameters were found between boys and girls. Results based on a Partial Credit Model indicate an absence of DIF in eight out of nine items. Consistent with other studies in attention deficit hyperactivity disorder (ADHD) our results imply that the same level of rating for a symptom has the same meaning for boys and girls.

Keywords. DIFFERENTIAL ITEM FUNCTIONING, PARTIAL CREDIT MODEL, ATTENTION DEFICIT, GENDER

1. Introduction

Differential item functioning (DIF) is a concept originating from psychometrics describing the statistical phenomenon that items are measuring the same abilities differently for members of separate subgroups (e.g., gender, culture). The detection of DIF indicates an unexpected item behavior and seriously jeopardizes the interpretation of group differences [1]. DIF analysis mainly occurs in the development of psychological assessment tools measuring some latent construct (e.g., intelligence, conceptual understanding, or knowledge) to produce valid and reliable scores. During this process, assessment tests are evaluated for differences in performance among subgroups to gather evidence of the presence or absence of a potential test bias [2].

In probabilistic test theory, the probability of a correct item response depends on both the characteristics of the item and the ability of the person. The functional relation between the two is expressed as an item response function and depicted in an item characteristic curve (ICC). When DIF is present, the ICCs of the same item differ for different subgroups. There are mainly two forms of DIF. Uniform DIF is present when

[1] Corresponding Author, Sebastian Appelbaum, Sebastian.Appelbaum@uni-wh.de.

one group is continuously superior across all levels of ability so that the ICCs are shifted but never cross. In contrast, in nonuniform or crossing DIF the ICCs intersect so that the direction of DIF changes at a certain point of ability.

While there is a vast amount of research and literature on DIF in the field of educational screening and career assessment, DIF analysis has hardly been applied in the field of clinical assessment. Only in the last decades, DIF has been evaluated for questionnaires related to disordered eating [3], the Hospital Anxiety and Depression Scale used in patients with dermatitis [4], the Geriatric Depression Scale [5] or parent ratings of ADHD symptoms [6].

Especially the latter case is a highly relevant topic. ADHD is one of the most common neurodevelopmental disorders mainly occurring in childhood with the highest prevalence estimates of 4-7% in western countries [7]. Data also suggests that there is an influence of the child's gender on the prevalence 2% girls, 6% boys, which has already been subject to a DIF analysis [7, 8].

This paper aims at analyzing the presence of gender related DIF in data from an already existing cross-sectional multicenter study (Approval from the Ethical committee of the Witten/Herdecke University No. 33/2012) of 1449 children from 14 pediatric wards assessed by a structured questionnaire on ADHD by their parents. After introducing the Partial Credit Model (PCM), the model is applied and its implications for health services research are critically discussed.

2. Methods

At the regular medical screening at the age of two years (U7) parents were informed about the study and gave consent to participate. They were asked to complete a questionnaire on the behavior of their child including sociodemographic data and an ADHS screening. All screening items used an ordinal five-point Likert scale ranging from 1 ="strongly disagree" to 5= "strongly agree" with an additional class for "unknown" treated as "missing". From the screening battery (56 items), five items on hyperactivity and four items on attention deficits were preselected by a principle component analysis with Varimax rotation (data not shown).

A PCM was applied to the response data. To estimate the person parameters θ_p and the item parameters β_{ij} we used the following response category probability characterization:

$$\pi_{pix} = P\left(x_{pi} = x \,|\theta_p, \beta_{ik}\right) = \frac{\exp[\sum_{j=0}^{x}(\theta_p - \beta_{ij})]}{\sum_{k=0}^{m_i}\exp[\sum_{j=0}^{k}(\theta_p - \beta_{ij})]} \quad (1)$$

In this equation, π_{pix} is the conditional probability of selecting category x from the m_i categories for an item i, θ_p is the capability level of the parents and β_{ij} is the category threshold parameters.

To investigate DIF within this model, we applied three proportional odds logistic regression models for each item. Item response is the dependent variable and the independent variables are (i) θ_p only, (ii) θ_p + gender and (iii) θ_p + gender + interaction. Likelihood ratio tests (LRT) between model (i) and model (iii) were applied to detect DIF, while LRTs between model (i) and model (ii) were used to detect uniform DIF (U-DIF). For all calculations the R package TAM was used [9]. To

correct for multiple testing, we applied a Bonferroni correction and tested at a nominal level of $\alpha_{corr} = 0.05 / 18 = 0.00278$.

3. Results

A total of 1449 children (mean age: 1.94 ± 0.14 years; 51.2% male) were included. No significant variations in age, parental education, family status and number of siblings were found between male and female children (see Table 1 for sociodemographic data).

Table 1. Sociodemographic and item characteristics

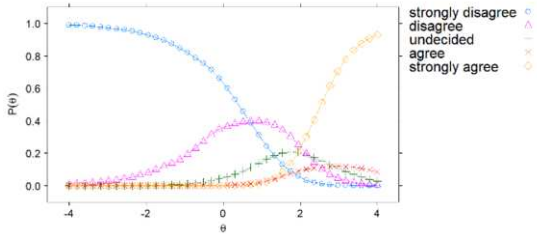
	Male (n=742)	Female (n=707)	Total (N=1449)
Age			
Mean ± SD	1.94 ± 0.15	1.94 ± 0.13	1.94 ± 0.14
Median [IQR]	1.92 [0.08]	1.92 [0.08]	1.92 [0.08]
Highest parental education			
Lower secondary school	40 (4.2)	29 (5.4)	69 (4.8)
Secondary school	102 (16.8)	117 (13.9)	219 (15.3)
A-Level	68 (10.2)	71 (9.3)	139 (9.7)
University degree	240 (32.0)	223 (32.7)	463 (32.4)
Completed vocational training	283 (36.6)	255 (38.5)	538 (37.6)
Other	2 (0.2)	1 (0.2)	3 (0.2)
Family status			
Single	91 (12.3%)	100 (14.4%)	191 (13.3%)
Married	584 (79.2%)	539 (77.4%)	1123 (78.4%)
Divorced	34 (4.6%)	35 (5.0%)	69 (4.8%)
Other	27 (3.9%)	22 (3.2%)	49 (3.5%)
Number of siblings			
0	329 (44.6%)	273 (39.0%)	602 (41.9%)
1	277 (37.5%)	298 (42.6%)	575 (40.0%)
2	98 (13.3%)	92 (13,1%)	190 (13.2%)
3 and more	34 (4.6%)	37 (5.3%)	71 (4.9%)

Table 2. Results of the PCM: Item difficulty and DIF detection statistics

Item	Item-Difficulty	DIF LRT.	DIF *p*	U-DIF LRT.	U-DIF *p*
1. My child quickly gets excited about something, but then loses interest	0.28	0.42	0.81	0.41	0.52
2. My child can be distracted quickly	-0.05	1.36	0.51	0.23	0.63
3. My child is restless, fidgety, hectic	1.25	0.27	0.87	0.09	0.77
4. My child is constantly on the move	-1.02	3.38	0.18	2.37	0.12
5. My child is often silly and hyped	0.91	1.77	0.41	0.41	0.52
6. My child screams often and intensely without quieting down	1.70	10.71	0.0047	10.38	**0.0013**
7. My child needs to be constantly driven	1.87	7.11	0.03	0.00	0.99
8. My child is interrupting or constantly disturbing others	1.41	0.30	0.89	0.03	0.86
9. My child always wants to be in the focus	0.77	4.98	0.08	0.01	0.93

As can be seen in Table 2, only one of the items was found to exhibit DIF. Specifically, Item 6 showed nonuniform DIF ($\chi^2 = 10.71$, $df = 2$, $p = 0.0047$). This effect mainly results from a shift in the difficulty parameter or uniform DIF ($\chi^2 = 10.38$, $df = 1$, $p = 0.0013$). The ICCs of this item are depicted in Figure 1 for boys and girls separately. It is obvious that the DIF in this item is mainly driven by the two highest categories "strongly agree" and "agree". Parents of boys that are high on latent ADHD tend to check "strongly agree" more likely than "agree". Parents of girls with the same ADHD characteristic prefer to check "agree" much more than "strongly agree".

(a) boys



(b) girls

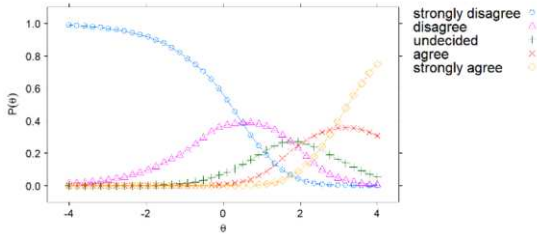


Figure 1: ICC plots by gender for Item 6 "My child screams often and intensely without quieting down"

4. Discussion

The present study applied a PCM to nine items of an ADHD screening test and then used LRTs to test for DIF as a function of gender. Unlike methods of classical test theory, such as analysis of variance or item discrimination, the probabilistic approach allows disentangling item and person parameters and hence detecting DIF at all [10]. DIF-testing thus is very useful to establish empirical evidence for measurement equivalence in health-related assessments.

Our results indicate an absence of DIF in eight out of nine items, which implies that given the same level of symptomatology, parents of boys and girls use to respond on the screening questionnaire the same way on most of the items. The exception is one item on "screaming often and intensely", where parents of boys tend to be more acquiescent as compared to the parents of girls. This tendency is compatible with the prevalent gender stereotype that allows boys more than girls to show this kind of behavior. Taken together, the ratings of both boys and girls tend to be based "on the same amount of the trait in each group" [6]. Thus, comparable score values reflect a comparable amount of hyperactivity or attention deficits across gender. To what extent the bias on the only item with DIF is responsible for gender differences in ADHD

remains an open question for future research. Moreover our analysis might also still be predisposed to making a Type I error due to multiple testing although Bonferroni correction was applied.

Our findings however are consistent with other studies in ADHD and support the idea of gender equivalence in the parental ratings of ADHD. Gender differences in the prevalence of ADHD hence are unlikely to be caused by differential reporting of the symptoms in boys and girls.

5. Conclusion

While DIF i.e. in educational psychology is mainly used for dichotomous items, health services research is much more complex with polytomous categories like in our study. And although there are some promising applications, there is still an urgent need for transferring psychometric concepts into biometrical research and vice versa.

6. Conflict of Interest

The authors state that they have no conflict of interests.

7. Acknowledgement

We would like to thank all participating parents, children and physicians in this survey.

References

[1] P. Martinková,, A. Drabinová, Y. L. Liaw, E. A. Sanders, J. L. McFarland, R. M. Price, Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education* **16(2)**, (2017), rm2.

[2] M. De Beer, Use of differential item functioning (DIF) analysis for bias analysis in test construction. *SA Journal of Industrial Psychology,* **30**, (2004), 52-58

[3] D. Thielemann, F. Richter, B. Strauss, E. Braehler, U. Altmann, U. Berger, Differential item functioning in brief instruments of disordered eating. *European Journal of Psychological Assessment,* (2018), Accepted

[4] J. I. Silverberg, J. M. Gelfand, D. J. Margolis, M. Boguniewicz, L. Fonacier, M. H. Grayson, E. L. Simpson, Measurement properties of Hospital Anxiety and Depression Scale used in atopic dermatitis in adults. *Journal of Investigative Dermatology*, (2018), Accepted.

[5] F. Chiesi, C. Primi, M. Pigliautile, M. Baroni, S. Ercolani, V. Boccardi, C. Ruggiero, P. Mecocci, Is the 15-item Geriatric Depression Scale a fair screening tool? A differential item functioning analysis across gender and age. *Psychological Reports* **121(6)**, (2018), 1167-1182.

[6] R. Gomez, Parent Ratings of ADHD symptoms: Generalized partial credit model analysis of differential item functioning across gender. *Journal of Attention Disorders,* **16**, (2012), 276-283.

[7] M. K. Akmatov, A. Steffen, J. Holstiege, R. Hering, M. Schulz, J. Bätzing, Trends and regional variations in the administrative prevalence of attention-deficit/hyperactivity disorder among children and adolescents in Germany. *Scientific reports*, **8(1)**, (2018),17029.

[8] R. Gomez, Testing gender differential item functioning for ordinal and binary scored parent rated ADHD symptoms. *Personality and Individual Differences* **42(4),** (2007), 733-742.

[9] A. Robitzsch, T. Kiefer, M. Wu. TAM: Test analysis modules. R package version, 2-0. (2017)

[10] M. M. Langer, C. D. Hill, D. Thissen, T. M. Burwinkle, J.W. Varni, D.A. DeWalt. Item response theory detected differential item functioning between healthy and ill children in quality-of-life measures. *Journal of Clinical Epidemiology*, **61(3)**, (2008), 268-276.